



TITLE Schnarch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 2204)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (21-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNNL)

DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbiology.org>  
contact: amadan@systemsbiology.org

Anup Madan, Jessica Fane, Erin Helton, Mark Kettman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNNL at: <http://image.llnl.gov>

Series: TRAK Plate: 47 Row: 1 Column: 8  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 14210481.  
Differences found between this sequence and the human genome (build  
35) are described in misc\_difference features below and these  
differences were also compared to chimpanzee genomic sequences  
available as of Sep 03, 03.

FEATURES

source  
1..2204  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:32924 IMAGE:5267610"  
/tissue\_type="Testis"

```
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
1..2204
/gene="HTPAP"
/db_xref="LocusID:84513"

misc_difference 1
/gene="HTPAP"
/note="1 base at the 5' end does not align to the human
genome."
5

misc_difference 5^6
/gene="HTPAP"
/note="G' in cDNA is 'C' in the human genome. The
chimpanzee genome agrees with the cDNA sequence,
suggesting that this difference is unlikely to be due to
an artifact."
205..876
/CDS
/gene="HTPAP"
/codon_start=1
/product="HTPAP protein"
/protein_id="AAH33025.1"
/db_xref="GI:21542541"
/db_xref="LocusID:84513"
/translation="MWLYRNPPVEAIFYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRD
SRQACLAASLALANGVFTNTIKLIVGRPRPDFYRCFPDGLAHSIDLMTGDKDVNE
GRKSFPSGHSSFAFAGLAFASFYFLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALS
RTCDYKHHWQDVLVGSMIGMTFAYVCYRQYYPPPLTDAECHKPFQDKLVLSTAQKPGDS
YCFDI"
379..786
/u
/gene="HTPAP"
/note="PAP2; Region: PAP2 superfamily. This family
includes the enzyme type 2 phosphatidic acid phosphatase
(PAP2), Glucose-6-phosphatase EC:3.1.3.9,
Phosphatidylglycerophosphatase B EC:3.1.3.27 and bacterial
acid phosphatase EC:3.1.3.2"
/db_xref="CDD:pfam01569"
```

misc\_difference 1169..1170  
 /gene="HTPAP"  
 /note="2 bases in cDNA are not found in the human genome.  
 The chimpanzee genome agrees with the cDNA sequence,  
 suggesting that this difference is unlikely to be due to  
 an artifact."

misc\_difference 2188..2204  
 /gene="HTPAP"  
 /note="polyA tail: 17 bases do not align to the human  
 genome."

ORIGIN	Sequence
1	agcgggttcg
61	ccaggcaggc
121	ggcggacctgc
181	cagagactca
241	tatttccca
301	ctggccaaat
361	agcccttgc
421	ccacgc
481	tgtacagg
541	tcctttgc
601	ttcacacc
661	tttgca
721	ctagttgg
781	cctctgact
841	cagaaggct
901	ggctcatg
961	ggaccagg
1021	agtgtgtc
1081	ggagctgg
1141	ttagaccc
1201	ctagcc
1261	aagg
1321	gc
1381	ctgt
1441	atct
1501	caca
1561	ttaatt
1621	tttgatt
1681	aatcttag
1741	ctgtgc
1801	ataaa
1	cgcccttc
61	gggggggg
121	tttttccc
181	ggatgtgg
241	gttgtttt
301	ttctcaaa
361	tggcttca
421	ttttctcta
481	tttgcgtt
541	tttgcgtt
601	tttttttt
661	tttttttt
721	tttttttt
781	tttttttt
841	tttttttt
901	tttttttt
961	tttttttt
1021	tttttttt
1081	tttttttt
1141	tttttttt
1201	tttttttt
1261	tttttttt
1321	tttttttt
1381	tttttttt
1441	tttttttt
1501	tttttttt
1561	tttttttt
1621	tttttttt
1681	tttttttt
1741	tttttttt
1801	tttttttt
1	ggggcggcc
61	gggggggg
121	gggggggg
181	gggggggg
241	gggggggg
301	gggggggg
361	gggggggg
421	gggggggg
481	gggggggg
541	gggggggg
601	gggggggg
661	gggggggg
721	gggggggg
781	gggggggg
841	gggggggg
901	gggggggg
961	gggggggg
1021	gggggggg
1081	gggggggg
1141	gggggggg
1201	gggggggg
1261	gggggggg
1321	gggggggg
1381	gggggggg
1441	gggggggg
1501	gggggggg
1561	gggggggg
1621	gggggggg
1681	gggggggg
1741	gggggggg
1801	gggggggg
1	ggggcggcc
61	ggggcggcc
121	ggggcggcc
181	ggggcggcc
241	ggggcggcc
301	ggggcggcc
361	ggggcggcc
421	ggggcggcc
481	ggggcggcc
541	ggggcggcc
601	ggggcggcc
661	ggggcggcc
721	ggggcggcc
781	ggggcggcc
841	ggggcggcc
901	ggggcggcc
961	ggggcggcc
1021	ggggcggcc
1081	ggggcggcc
1141	ggggcggcc
1201	ggggcggcc
1261	ggggcggcc
1321	ggggcggcc
1381	ggggcggcc
1441	ggggcggcc
1501	ggggcggcc
1561	ggggcggcc
1621	ggggcggcc
1681	ggggcggcc
1741	ggggcggcc
1801	ggggcggcc
1	ggggcggcc
61	ggggcggcc
121	ggggcggcc
181	ggggcggcc
241	ggggcggcc
301	ggggcggcc
361	ggggcggcc
421	ggggcggcc
481	ggggcggcc
541	ggggcggcc
601	ggggcggcc
661	ggggcggcc
721	ggggcggcc
781	ggggcggcc
841	ggggcggcc
901	ggggcggcc
961	ggggcggcc
1021	ggggcggcc
1081	ggggcggcc
1141	ggggcggcc
1201	ggggcggcc
1261	ggggcggcc
1321	ggggcggcc
1381	ggggcggcc
1441	ggggcggcc
1501	ggggcggcc
1561	ggggcggcc
1621	ggggcggcc
1681	ggggcggcc
1741	ggggcggcc
1801	ggggcggcc
1	ggggcggcc
61	ggggcggcc
121	ggggcggcc
181	ggggcggcc
241	ggggcggcc
301	ggggcggcc
361	ggggcggcc
421	ggggcggcc
481	ggggcggcc
541	ggggcggcc
601	ggggcggcc
661	ggggcggcc
721	ggggcggcc
781	ggggcggcc
841	ggggcggcc
901	ggggcggcc
961	ggggcggcc
1021	ggggcggcc
1081	ggggcggcc
1141	ggggcggcc
1201	ggggcggcc
1261	ggggcggcc
1321	ggggcggcc
1381	ggggcggcc
1441	ggggcggcc
1501	ggggcggcc
1561	ggggcggcc
1621	ggggcggcc
1681	ggggcggcc
1741	ggggcggcc
1801	ggggcggcc
1	ggggcggcc
61	ggggcggcc
121	ggggcggcc
181	ggggcggcc
241	ggggcggcc
301	ggggcggcc
361	ggggcggcc
421	ggggcggcc
481	ggggcggcc
541	ggggcggcc
601	ggggcggcc
661	ggggcggcc
721	ggggcggcc
781	ggggcggcc
841	ggggcggcc
901	ggggcggcc
961	ggggcggcc
1021	ggggcggcc
1081	ggggcggcc
1141	ggggcggcc
1201	ggggcggcc
1261	ggggcggcc
1321	ggggcggcc
1381	ggggcggcc
1441	ggggcggcc
1501	ggggcggcc
1561	ggggcggcc
1621	ggggcggcc
1681	ggggcggcc
1741	ggggcggcc
1801	ggggcggcc

1861 ctcatttgac cataactgta cttgtgagct tctttccccc cattcatgac tgaagatctg  
1921 ctcAAatgca ccaacactgc caagtgacta aggttagaaaa gaaaaatgac aggtatgc  
1981 atctgaagga cagatgaaatc ttttttgc ctttttgc aatggaaat aagggaaacaaat  
2041 tatggatgt catcagaatg gatgccatag gacctacagg tccctttctc tttatttga  
2101 ttatacttta aatatgacat tgtttttat gtgtatgttc ctatatttc aatgtatctt  
2161 ttcccttcag taaaacctgat attcaaaacaa aaaaaaaa aaaa  
//

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)



## Sequence Revision History

PubMed	Nucleotide	Protein	Genome	Structure	PMC	Taxonomy	OMIM	Books
<a href="#">Find (Accessions, GI numbers or Fasta style Seq/ids)</a>	BC033025							
<a href="#">About Entrez</a>	<input checked="" type="checkbox"/> Show difference in GenBank/GenPept Diff	<input type="checkbox"/> format	<a href="#">Show</a>	<a href="#">Update Date</a>	<a href="#">Oct 6 2003 7:25 PM</a>			
<a href="#">Entrez</a>	<b>Gi</b> 21542540	<b>Version</b> 1						
<a href="#">Search for Genes</a>	<b>Gi</b> 21542540	<b>Version</b> 1						

Search for Genes  
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

### Help|FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

### Differences in LOCUS names:

old: "LOCUS	BC033025	2204	bp	mRNA	linear	PRI
24-JUN-2002"	^					
new: "LOCUS	BC033025	2204	bp	mRNA	linear	PRI
06-OCT-2003"	^					

### Check sequence revision history

How to create WWW links to Entrez

LinkOut  
Cubby

### Differences in DEFINITION:

old: "Homo sapiens, Similar to HTPAP protein, clone MGC:32924 IMAGE:5267610,	^
mRNA, complete cds."	
new: "Homo sapiens HTPAP protein, mRNA (cDNA clone MGC:32924 IMAGE:5267610),	^
complete cds."	

### Differences in REFERENCE:

Related resources	Changed reference:
BLAST	refline: old: "1 (bases 1 to 2204)" new: "2 (bases 1 to 2204)" ^
Reference sequence project	
LocusLink	New reference:

```

refline: "1 (bases 1 to 2204)"
authors: "Strausberg, R.L., Feingold, E.A., Grouse, L.H.,  

          Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.,  

          Shenmen, C.M., Schuler, G.D., Altschul, S.F.,  

          Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  

          Hopkins, R.F., Jordan, H., Moore, T., Max, S.I.,  

          Wang, J., Hsieh, F., Diatchenko, L., Marusina, K.,  

          Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M.,  

          Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  

          Scheetz, T.E., Brownstein, M.J., Usdin, T.B.,  

          Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S.,  

          Loquellano, N.A., Peters, G.J., Abramson, R.D.,  

          Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  

          McKernan, K.J., Malek, J.A., Gunaratne, P.H.,  

          Richards, S., Worley, K.C., Hale, S., Garcia, A.M.,  

          Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M.,  

          Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J.,  

          Helton, E., Kettman, M., Madan, A., Rodrigues, S.,  

          Sanchez, A., Whiting, M., Madan, A., Young, A.C.,  

          Shevchenko, Y., Bouffard, G.G., Blakesley, R.W.,  

          Touchman, J.W., Green, E.D., Dickson, M.C.,  

          Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  

          Butterfield, Y.S., Krzywinski, M.I., Skalska, U.,  

          Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J.  

          and Marra, M.A."
title: "Generation and initial analysis of more than 15,000  

        full-length human and mouse cDNA sequences"
journal: "Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903  

          (2002)"
muid: "22388257"
pmid: "12477932"

```

## Differences in FEATURE.CDS:

```

-----  

Changed feature CDS 205..876  

Changed /product =  

  old: "Similar to HTPAP protein"  

          ^  

  new: "HTPAP protein"  

          ^  

New /db_xref = "LocusID:84513"  

  New /gene = "HTPAP"

```

Differences in FEATURE.gene:

```
-----  
New feature  gene  1...2204  
/db_xref      = "LocusID:84513"  
/gene         = "HTPAP"
```

Differences in FEATURE.misc\_feature:

```
-----  
New feature  misc_feature  379...786  
/db_xref      = "CDD:pfam01569"  
/gene         = "HTPAP"  
/note         = "PAP2; Region: PAP2 superfamily. This  
family includes the enzyme type 2  
phosphatidic acid phosphatase (PAP2),  
Glucose-6-phosphatase EC:3.1.3.9,  
Phosphatidylglycerophosphatase B  
EC:3.1.3.27 and bacterial acid phosphatase  
EC:3.1.3.2"
```

[Disclaimer](#) | [Write to the Help Desk](#)  
NCBI | NLM | NIH